

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 15, 2001, 11:11:12 ; Search time 16.02 Seconds

(without alignments)
30.238 Million cell updates/sec

Title: US-09-288-719-3

Sequence: 1 GGGSGGGRSGGGGS 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 88757 segs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	70	86.4	465	FXD3_MOUSE	O61060 mus musculus
2	66	81.5	779	M130_STRPU	P08472 strongyloce
3	65	80.2	104	HOL3_HOIDI	Q25055 holotrichia
4	65	80.2	643	K2C1_HUMAN	P04264 homo sapien
5	63	77.8	206	TWST_MOUSE	P26687 mus musculu
6	63	77.8	214	GRP2_NICSY	P27484 nicotiana s
7	63	77.8	265	YPG1_ZYMO	P30794 zymomonas m
8	63	77.8	431	HXB3_HUMAN	P14651 homo sapien
9	63	77.8	433	HXB3_MOUSE	P09026 mus musculu
10	63	77.8	1434	KDGE_MOUSE	Q09103 drosophila
11	62	76.5	157	GRPA_MAIZE	P10979 zea mays (m
12	62	76.5	157	GRP_DAUCA	Q03878 daucus caro
13	62	76.5	165	GRP1_ORYSA	P25074 oryza sativ
14	62	76.5	168	GRP2_SORBI	Q09070 sorghum dic
15	62	76.5	280	CHIA_MAIZE	P29022 zea mays (m
16	62	76.5	321	PUR_MOUSE	P42669 mus musculu
17	62	76.5	322	PUR_HUMAN	Q00577 homo sapien
18	62	76.5	323	HXB_MOUSE	P23813 mus musculu
19	62	76.5	440	FXGA_CHICK	Q08937 gallus galli
20	62	76.5	495	BRNL_MOUSE	P31361 mus musculu
21	62	76.5	497	BRNL_RAT	Q61262 rattus norv
22	62	76.5	500	BRNL_HUMAN	P20264 homo sapien
23	62	76.5	539	DOP2_DROME	Q24563 drosophila
24	62	76.5	569	K1CJ_MOUSE	P02535 mus musculu
25	62	76.5	622	SR68_CANFA	Q00004 canis famli
26	62	76.5	663	DUS8_MOUSE	Q09112 mus musculu
27	62	76.5	688	BOMD_MOUSE	O54839 mus musculu
28	62	76.5	979	REF1_HUMAN	P22670 homo sapien
29	62	76.5	1627	TP2B_CHICK	O42131 gallus galli
30	61	75.3	316	LORI_HUMAN	P23490 homo sapien
31	60	74.1	266	CANS_RABIT	P06813 oryctolagus
32	60	74.1	332	STX3_HUMAN	O95343 homo sapien
33	60	74.1	401	CSP_PLACC	P08674 plasmodium

34	60	74.1	410	1	BRB3_HUMAN	Q12837 homo sapien
35	60	74.1	653	1	YKR2_CAEEL	P34308 caenorhabdi
36	60	74.1	722	1	COAT_PAVHE	P03136 hamster par
37	59	72.8	43	1	GRW1_LYCES	Q01157 lycopersico
38	59	72.8	151	1	GDF7_MOUSE	P43029 mus musculu
39	59	72.8	263	1	CANS_BOVIN	P13135 bos taurus
40	59	72.8	266	1	CANS_PIG	P04574 sus scrofa
41	59	72.8	268	1	CANS_HUMAN	P04632 homo sapien
42	59	72.8	268	1	EP34_HCVYA	P16768 human cytom
43	59	72.8	367	1	BET3_MESAV	O09029 mesocricetu
44	59	72.8	387	1	SOX1_HUMAN	O00570 homo sapien
45	59	72.8	391	1	SOX1_MOUSE	P53783 mus musculu

ALIGNMENTS

RESULT	1	FXD3_MOUSE	STANDARD:	PRT:	465 AA.
ID	FXD3_MOUSE				
AC	Q61060:				
DT	15-DEC-1998 (Rel. 37, Created)				
DT	15-DEC-1998 (Rel. 37, Last sequence update)				
DT	30-MAY-2000 (Rel. 39, Last annotation update)				
DE	FORKHEAD BOX PROTEIN D3 (HNF3/PH TRANSCRIPTION FACTOR GENESIS)				
DE	(HEPATOCYTE NUCLEAR FACTOR 3 FORKHEAD HOMOLOG 2) (HFH-2).				
GN	FOX D3 OR HFH2.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
RM	[1]				
RP	SEQUENCE FROM N.A.				
RA	Hromas R.A., Costa R.H., Xu D., Sutton J.L.;				
RL	Submitted (NOV-1995) to the EMBL/Genbank/DBJ databases.				
RM	[2]				
RP	SEQUENCE FROM N.A.				
RA	Labosky P.A., Kaestner K.H.;				
RT	"The winged helix transcription factor Hfh2 is expressed in neural crest and spinal cord during mouse development."				
RL	Submitted (MAY-1998) to the EMBL/Genbank/DBJ databases.				
CC	- FUNCTION: PROBABLE TRANSCRIPTION FACTOR.				
CC	- SUBCELLULAR LOCATION: NUCLEAR.				
CC	- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).				
CC	-----				
DR	EMBL: U41047; AA87569.1; -				
DR	EMBL: AF067421; AAC28352.1; -				
DR	MGD: MGI:1347473; FOXD3.				
DR	INTERPRO: IPR001766; -				
DR	PFAM: PF00250; Fork_head. 1.				
DR	PRINTS: PR00053; FORKHEAD.				
DR	PROSITE: PS00657; FORK_HEAD_1; 1.				
DR	PROSITE: PS00658; FORK_HEAD_2; 1.				
DR	PROSITE: PS50039; FORK_HEAD_3; 1.				
KW	DNA-binding; Nuclear protein; Transcription regulation.				
FT	DOMAIN 106 113				
FT	DNA_BIND 131 225				
FT	DOMAIN 252 257				
FT	DNA_BIND 265 270				
FT	DOMAIN 275 281				
FT	DNA_BIND 380 393				
FT	DOMAIN 385 399				
FT	DNA_BIND 447 457				
FT	DOMAIN 455 457				
SO	SEQUENCE 465 AA; 47092 MW; 6F8B5B3D8C7564D CRC64;				

	Query Match	86.4%;	Score 70;	DB 1;	Length 465;
	Best Local Similarity	86.7%;	Pred. No. 0.52;		
Matches	13; Conservative	0;	Mismatches	2;	Indels 0; Gaps 0;
OY	1 GGCGGCGRAGCGGS	15			
		1			
Dd	380 GGCGGCGRAGCGGS	394			

AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT	AV	AW	AX	AY	AZ	BA	BB	BC	BD	BE	BF	BG	BH	BI	BJ	BK	BL	BM	BN	BO	BP	BQ	BR	BS	BT	BV	BW	BX	BY	BZ	CA	CB	CC	CD	CE	CF	CG	CH	CI	CJ	CK	CL	CM	CN	CO	CP	CQ	CR	CS	CT	CU	CV	CW	CX	CY	CZ	DA	DB	DC	DD	DE	DF	DG	DH	DI	DJ	DK	DL	DM	DN	DO	DP	DQ	DR	DS	DT	DU	DV	DW	DX	DY	DZ	EA	EB	EC	ED	EE	EF	EG	EH	EI	EJ	EK	EL	EM	EN	EO	EP	EQ	ER	ES	ET	EU	EV	EW	EX	EY	EZ	FA	FB	FC	FD	FE	FF	FG	FH	FI	FJ	FK	FL	FM	FN	FO	FP	FQ	FR	FS	FT	FU	FV	FW	FX	FY	FZ	GA	GB	GC	GD	GE	GF	GG	GH	GI	GJ	GK	GL	GM	GN	GO	GP	GQ	GR	GS	GT	GU	GV	GW	GX	GY	GZ	HA	HB	HC	HD	HE	HF	HG	HH	HI	HJ	HK	HL	HM	HN	HO	HP	HQ	HR	HS	HT	HU	HV	HW	HX	HY	HZ	IA	IB	IC	ID	IE	IF	IG	IH	II	IJ	IK	IL	IM	IN	IO	IP	IQ	IR	IS	IT	IU	IV	IW	IX	IY	IZ	JA	JB	JC	JD	JE	JF	JG	JH	JI	JJ	JK	JL	JM	JN	JO	JP	JQ	JR	JS	JT	JU	JV	JW	JX	JY	JZ	KA	KB	KC	KD	KE	KF	KG	KH	KI	KJ	KK	KL	KM	KN	KO	KP	KQ	KR	KS	KT	KU	KV	KW	KX	KY	KZ	LA	LB	LC	LD	LE	LF	LG	LH	LI	LJ	LK	LL	LM	LN	LO	LP	LQ	LR	LS	LT	LU	LV	LW	LX	LY	LZ	MA	MB	MC	MD	ME	MF	MG	MH	MI	MJ	MK	ML	MM	MN	MO	MP	MQ	MR	MS	MT	MU	MV	MW	MX	MY	MZ	NA	NB	NC	ND	NE	NF	NG	NH	NI	NJ	NK	NL	NM	NO	NP	NQ	NR	NS	NT	NU	NV	NW	NX	NY	NZ	OA	OB	OC	OD	OE	OF	OG	OH	OI	OJ	OK	OL	OM	ON	OO	OP	OQ	OR	OS	OT	OU	OV	OW	OX	OY	OZ	PA	PB	PC	PD	PE	PF	PG	PH	PI	PJ	PK	PL	PM	PN	PO	PP	PQ	PR	PS	PT	PV	PW	PX	PY	PZ	QA	QB	QC	QD	QE	QF	QG	QH	QI	QJ	QK	QL	QM	QN	QO	QP	QQ	QR	QS	QT	QU	QV	QW	QX	QY	QZ	RA	RB	RC	RD	RE	RF	RG	RH	RI	RJ	RK	RL	RM	RN	RO	RP	RQ	RR	RS	RT	RU	RV	RW	RX	RY	RZ	SA	SB	SC	SD	SE	SF	SG	SH	SI	SJ	SK	SL	SM	SN	SO	SP	SQ	SR	SS	ST	SU	SV	SW	SX	SY	SZ	TA	TB	TC	TD	TE	TF	TG	TH	TI	TJ	TK	TL	TM	TN	TO	TP	TQ	TR	TS	TT	TU	TV	TW	TX	TY	TZ	UA	UB	UC	UD	UE	UF	UG	UH	UI	UJ	UK	UL	UM	UN	UO	UP	UQ	UR	US	UT	UU	UV	UW	UX	UY	UZ	VA	VB	VC	VD	VE	VF	VG	VH	VI	VJ	VK	VL	VM	VN	VO	VP	VQ	VR	VS	VT	VU	VV	VW	VX	VY	VZ	WA	WB	WC	WD	WE	WF	WG	WH	WI	WJ	WK	WL	WM	WN	WO	WP	WQ	WR	WS	WT	WU	WV	WW	WX	WY	WZ	XA	XB	XC	XD	XE	XF	YG	YH	YI	YJ	YK	YL	YM	YN	YO	YP	YQ	YR	YS	YT	YU	YV	YW	YX	YY	YZ	ZA	ZB	ZC	ZD	ZE	ZF	ZG	ZH	ZI	ZJ	ZK	ZL	ZM	ZN	ZO	ZP	ZQ	ZR	ZS	ZT	ZU	ZV	ZW	ZX	ZY	ZZ
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

```
QY . 1 GGGSGGRASGGCG 14
      |||:|||| |||
Db 57 GGGGAGGRGGGGCG 70
```

ID	HOLD1	STANDARD	PRT	104 AA.
AC	025055			
DT	01-NOV-1997	(Rel. 35, Created)		
DT	01-NOV-1997	(Rel. 35, last sequence update)		
DT	01-NOV-1997	(Rel. 35, last annotation update)		
DE	HOLOTRICIN 3 PRECURSOR.			
OS	Holotrichia diomphalia.			
OC	Eukaryota, Metazoa, Arthropoda, Tracheata, Hexapoda, Insecta;			
OC	Pterygota, Neoptera, Endopterygota, Coleoptera, Polyphaga;			
OC	Scarabaeiformia; Scarabaeidae; Melolonthinae; Holotrichia.			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 21-40.			
RC	TISSUE-LARVAL HEMOLYMPH.			
RX	MEDLINE=96073722; Pubmed=853593.			
RA	Lee S.Y., Moon H.J., Kurata S., Natori S., Lee B.L.;			
RT	"Purification and cDNA cloning of an antifungal protein from the			
RT	hemolymph of Holotrichia diomphalia larvae."			
RL	Biol. Pharm. Bull. 18:1049-1052(1995).			
CC	-1- FUNCTION: HOLOTRICIN 3 HAS ANTIFUNGAL ACTIVITY AGAINST			
CC	C. ALBICANS.			
CC	-1- SIMILARITY: TO TENECIN 3.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL: D13744; BAA02889.1; -			
DR	HSSP; P30129; 4DPV.			
KW	Insect Immunity; Antibiotic; Hemolymph; Fungicide; Signal; Repeat.			
FT	SIGNAL	1	20	HOLOTRICIN 3.
FT	PEPTIDE	21	104	18 X 4 AA APPROXIMATE TANDEM REPEATS OF
FT	DOMAIN	27	98	H-G-G-G.
FT	REPEAT	27	30	1.
FT	REPEAT	31	34	2.
FT	REPEAT	35	38	3.
FT	REPEAT	39	42	4.
FT	REPEAT	43	46	5.
FT	REPEAT	47	50	6.
FT	REPEAT	51	54	7.
FT	REPEAT	55	58	8.
FT	REPEAT	59	62	9.
FT	REPEAT	63	66	10.
FT	REPEAT	67	70	11.
FT	REPEAT	71	74	12.
FT	REPEAT	75	78	13.
FT	REPEAT	79	82	14.
FT	REPEAT	83	86	15.
FT	REPEAT	87	90	16.
FT	REPEAT	91	94	17.
FT	REPEAT	95	98	18.
SO	SEQUENCE	104 AA;	9026 MW;	2799D681BFDCC725 CRC64;
QY	1	GGGGGCGRAGGGGS	15	80.2%; Score 65; DB 1; Length 104;
DB	64	GGGGGCGRAGGGGS	78	Best Local Similarity 80.0%; Pred. No. 0.46;
				Matches 12; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

RCCL_HUMAN	4	RESULT
ID	KCCL_HUMAN	STANDARD; PRT; 643 AA.
AC	P04264;	
DT	20-MAR-1987 (Rel. 04, Created)	
DT	01-FEB-1996 (Rel. 33, Last sequence update)	
DT	01-OCT-2000 (Rel. 40, Last annotation update)	
DE	KEATIN, TYPE II CYTOSKELETAL 1 (CYTOKERATIN 1) (K1) (CK 1) (67 KDA CYTOKERATIN) (HAIR ALPHA PROTEIN).	
DE	KRT1 OR KRTA.	
OC	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
RN	[1]	
RX	SEQUENCE FROM N.A.	
RX	MEDLINE=85166239; PubMed=2580302;	
RL	Johnson L.D., Idler W.W., Zhou X.-M., Roop D.R., Steinert P.M.;	
RL	"Structure of a gene for the human epidermal 67-kDa keratin.";	
RL	Proc. Natl. Acad. Sci. U.S.A. 82:1896-1900(1985).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RP	Whitlock N.V., Eady R.A., McGrath J.A.;	
RL	"Genomic organization of the human keratin 1 gene.";	
RL	Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.	
RN	[3]	
RX	PRELIMINARY SEQUENCE OF 151-643 FROM N.A.	
RX	MEDLINE=85207740; PubMed=2581954;	
RA	Steinert P.M., Parry D.A.D., Idler W.W., Johnson L.D., Steven A.C.,	
RA	Roop D.R.;	
RT	"Amino acid sequences of mouse and human epidermal type II keratins	
RT	of Mr. 67,000 provide a systematic basis for the structural and	
RT	functional diversity of the end domains of keratin intermediate	
RT	filament subunits.";	
RL	J. Biol. Chem. 260:7142-7149(1985).	
RN	[4]	
RP	REVISIONS, AND VARIANT EHK PRO-160.	
RX	MEDLINE=92386601; PubMed=1361288;	
RA	Chipev C.C., Korge B.P., Markova N., Bale S.J., Diglova J.J.,	
RT	Compton J.G., Steinert P.M.;	
RT	"A leucine-->proline mutation in the H1 subdomain of keratin 1	
RL	causes epidermolytic hyperkeratosis.";	
RL	Cell 70:821-828(1992).	
RN	[5]	
RP	VARIANTS EHK GLY-154; SER-187 AND PRO-192.	
RX	MEDLINE=94117869; PubMed=7507151;	
RA	Yang J.-M., Chipev C.C., Diglova J.J., Bale S.J., Marekov L.N.,	
RA	Steinert P.M., Compton J.G.;	
RT	"Mutations in the H1 and IA domains in the keratin 1 gene in	
RT	epidermolytic hyperkeratosis.";	
RL	J. Invest. Dermatol. 102:17-23(1994).	
RN	[6]	
RP	VARIANTS EHK PRO-185 AND SER-187.	
RX	MEDLINE=94117870; PubMed=7507152;	
RA	McLean W.H.I., Eady R.A., Dopping-Hepenstal P.J., McMillan J.R.,	
RA	Leigh I.M., Navsaria H.A., Higgins C., Harper J.I., Paige D.G.,	
RA	Morley S.M.;	
RT	"Mutations in the rod IA domain of keratins 1 and 10 in bullous	
RT	congenital ichthyosiform erythroderma (BCIE).";	
RL	J. Invest. Dermatol. 102:24-30(1994).	
RN	[7]	
RP	VARIANT EHK GLN-489.	
RX	MEDLINE=92376531; PubMed=1380725;	
RA	Rochengel J.A., Dominey A.M., Dempsey L.D., Longley M.A.,	
RA	Greenhalgh D.A., Gagne T.A., Huber M., Frenk E., Hohl D., Roop D.R.	
RT	"Mutations in the rod domains of keratins 1 and 10 in epidermolytic	
RT	hyperkeratosis.";	
RL	Science 257:1128-1130(1992).	
RN	[8]	
RP	VARIANT ALLELE 1B.	
RX	MEDLINE=93107743; PubMed=1281859;	
RA	Korge B.P., Compton J.G., Steinert P.M., Mische D.;	

RT		"The two size alleles of human keratin 1n are due to a deletion in the
Rt		ylic-rich carboxyl-terminal V2 subdomain.",
RL	J. Invest. Dermatol.	99:697-702(1992).
CC	-I-	SUBUNIT: HETERODIMER OF TWO TYPE I AND TWO TYPE II KERATINS.
CC	-I-	KERATIN I IS GENERALLY ASSOCIATED WITH KERATIN 10.
CC	-I-	TISSUE SPECIFICITY: THE SOURCE OF THIS PROTEIN IS NEONATAL
CC	-I-	FOOSKIN. THE 67-KDA TYPE II KERATINS ARE EXPRESSED IN TERMINALLY
CC	-I-	DIFFERENTIATING EPIDERMIS.
CC	-I-	POLYMORPHISM: THERE ARE TWO SIZE VARIANTS OF KRT1, TERMED 1A AND
CC	-I-	1B WITH ALLELIC FREQUENCIES OF 0.61 AND 0.39. 1B LACKS 7 LACKS
CC	-I-	7 RESIDUES COMPARED TO 1A.
CC	-I-	DISEASE: DEFECTS IN KRT10 AND KRT1 ARE THE CAUSE OF EPIDERMOLYTIC
CC	-I-	HYPERTROPHIC ERYTHRODERMA (EHE) (ALSO KNOWN AS BULLOUS CONGENITAL
CC	-I-	ICHTHYOSIFORM ERYTHRODERMA (BIE)); A HEREDITARY SKIN DISORDER
CC	-I-	CHARACTERIZED BY BLISTERING AND A MARKED THICKENING OF THE STRATUM
CC	-I-	CORNELM.
CC	-I-	MISCELLANEOUS: THERE ARE TWO TYPES OF CYTOSKELETAL AND
CC	-I-	MICROFILAMENTAR KERATIN: I (ACIDIC; 40-55 KDA) [K9 TO K20] AND II
CC	-I-	(NEUTRAL TO BASIC; 56-70 KDA) [K1 TO K8].
CC	-I-	SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC	-I-	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	-I-	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	-I-	the European Bioinformatics Institute. There are no restrictions on its
CC	-I-	use by non-profit institutions as long as its content is in no way
CC	-I-	modified and this statement is not removed. Usage by and for commercial
CC	-I-	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	-I-	or send an email to license@isb-sib.ch).
CC	-I-	-----
DR	EMBL;	AF237621; AA660327.1; -
DR	EMBL;	M98776; AAA47721.1; --
DR	PIR;	A02950; KRHD2.
DR	PIR;	A22940; A22940.
DR	HSSP;	P05412; IFOS.
DR	AARHUS/GHENT-2DPAGE;	4606; NEEHGE.
DR	MM;	139350; --
DR	MM;	113800; --
DR	INTERPRO;	IIPR001664; -
DR	INTERPRO;	IIPR003054; -
DR	PFAM;	PF00038; filament_1.
DR	PRINTS;	PR01276; TYPE2KERATIN.
DR	PROSITE;	PS00226; IF, 1.
KW	Intermediate filament; Coiled coil; Heptad repeat pattern; Keratin;	
KW	Disease mutation; Polymorphism; Phosphorylation.	
FT	INIT_MET	0
FT	DOMAIN	1 178 HEAD.
FT	DOMAIN	179 488 ROD.
FT	DOMAIN	489 643 TAIL.
FT	DOMAIN	179 214 COIL 1A.
FT	DOMAIN	215 233 LINER 1.
FT	DOMAIN	234 325 COIL 1B.
FT	DOMAIN	326 349 LINKER 12.
FT	DOMAIN	350 488 COIL 2.
FT	SITE	432 432 STUTTER.
FT	DOMAIN	1 150 GLY/PHE/SER-RICH.
FT	DOMAIN	501 640 GLY/SER-RICH.
FT	MOD_RES	65 65 PHOSPHORYLATION (BY SIMILARITY).
FT	VARIANT	154 154 /FtId=VAR_003853.
FT	VARIANT	160 160 L->P (IN EHK).
FT	VARIANT	185 185 /FtId=VAR_003854.
FT	VARIANT	187 187 S->P (IN EHK).
FT	VARIANT	192 192 /FtId=VAR_003855.
FT	VARIANT	192 192 S->S (IN EHK).
FT	VARIANT	192 192 /FtId=VAR_003856.
FT	VARIANT	311 311 S->P (IN EHK).
FT	VARIANT	311 311 I->V.
FT	VARIANT	329 329 /FtId=VAR_003858.
FT	VARIANT	329 329 I->T.
FT	VARIANT	357 357 /FtId=VAR_003859.
FT	VARIANT	357 357 N->Y.
FT	VARIANT	/FtId=VAR_003860.

FT VARIANT 489 489 E -> O (IN EHK).
/FTid=VAR_003861.
FT VARIANT 536 536 G -> C.
/FTid=VAR_003862.
FT VARIANT 632 632 R -> K.
/FTid=VAR_003863.
FT VARIANT 559 565 MISSING (IN ALLELE 1B).
/FTid=VAR_003864.
SQ SEQUENCE 643 AA; 65886 MM; DF945DC46257850 CRC64;

Query Match 80.2%; Score 65; DB 1; Length 643;
Best Local Similarity 80.0%; Pred. No. 2.2;
Matches 12; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 GGGGSGGRASGGGGS 15
||| |||| |||| |
Db 594 GGGSGGGRSGGGSS 608

RESULT 5
TWST_MOUSE STANDARD: PRT: 206 AA.
AC P26687; 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE TWIST RELATED PROTEIN (M-TWIST).
GN TWIST.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-91122450; PubMed-1840517;
RA Wolf C., Thisse C., Stoetzel C., Thisse B., Gerlinger P.,
RA Perrin-Schmitt F.;
RT "The M-twist gene of Mus is expressed in subsets of mesodermal cells
and is closely related to the Xenopus X-tw1 and the Drosophila twist
genes";
RL Dev. Biol. 143:363-373(1991).
RN [2]
RP FUNCTION.
RX MEDLINE-98001585; PubMed-9343420;
RA Hamamori Y., Wu H.Y., Sartorelli V., Kedes L.;
RT "The basic domain of myogenic basic helix-loop-helix (bHLH) proteins
is the novel target for direct inhibition by another bHLH protein,
Twist";
RL Mol. Cell. Biol. 17:6563-6573(1997).
CC -1- FUNCTION: PROBABLE TRANSCRIPTION FACTOR, WHICH SEEMS TO BE
INVOLVED IN THE NEGATIVE REGULATION OF CELLULAR DIFFERENTIATION AND
IN THE DIFFERENTIATION OF SEVERAL LINEAGES INCLUDING MYOGENESIS,
OSTEOGENESIS, AND NEUROGENESIS. INHIBITS MYOGENESIS BY
SEQUESTERING E PROTEINS, INHIBITING TRANS-ACTIVATION BY MYF2, AND
INTERACTING WITH MYO2 THROUGH PHYSICAL INTERACTION. THIS
INTERACTION PROBABLY INVOLVES THE BASIC DOMAINS OF BOTH PROTEINS.
CC -1- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
bHLH PROTEIN HOMODIMER.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- TISSUE SPECIFICITY: SUBSET OF MESODERMAL CELLS.
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (bHLH) FAMILY OF
TRANSCRIPTION FACTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M63649; AAA40514.1; -
DR EMBL; M63650; AAA40515.1; -
DR

DR HSSP: P10085; IMDY.
DR TRANSFAC; T01635; -
DR MGD; MGI:98872; TWIST.
DR INTERPRO: IPR001092; -
DR INTERPRO: IPR003015; -
DR PFAM: PF00010; HLH; 1.
DR PROSITE: PS00038; HELIX_LOOP_HELIX; 1.
KW Differentiation; Developmental protein; Nuclear protein; DNA-binding;
KM Transcription regulation.
FT DOMAIN 80 102 GLY-RICH.
FT DNA_BIND 112 124 BASIC DOMAIN.
FT DOMAIN 125 164 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT VARIANT 36 36 A -> R (IN CDNA).
FT VARIANT 91 91 G -> P (IN CDNA).
SQ SEQUENCE 206 AA; 21198 MM; 618AD8E9BE87C555 CRC64;

Query Match 77.8%; Score 63; DB 1; Length 206;
Best Local Similarity 73.3%; Pred. No. 1.3;
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 GGGGSGGRASGGGGS 15
||| |||| |||| |
Db 83 GGGGAGGGGGGGGGS 97

RESULT 6
GRP2_NICSY STANDARD: PRT: 214 AA.
ID GRP2_NICSY
AC P27484; 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE GLYCINE-RICH CELL WALL STRUCTURAL PROTEIN 2 PRECURSOR.
GN GRP-2.
OS Nicotiana sylvestris (Wood tobacco).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I;
OC Solanales; Solanaceae; Nicotiana.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-92003709; PubMed-1912512;
RA Obokata J., Ohme M., Hayashida N.;
RT "Nucleotide sequence of a cDNA clone encoding a putative glycine-rich
protein of 19.7 kDa in Nicotiana sylvestris";
RL Plant Mol. Biol. 17:953-955(1991).
CC -1- FUNCTION: RESPONSIBLE FOR PLASTICITY OF THE CELL WALL (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: CELL WALL (POTENTIAL).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X60007; CAA42622.1; -
DR PIR; S1731; KNT2S.
DR HSSP; P15277; IMC.
DR INTERPRO: IPR001878; -
DR INTERPRO: IPR002059; -
DR PFAM: PF00033; CSD; 1.
DR PFAM: PF00098; zf-CCHC; 2.
DR PRINTS; PRO0050; COLDSHOCK.
DR PRINTS; PRO0039; C2HCZNFINGER.
DR PROSITE; PS00352; COLD_SHOCK; UNKNOWN; 1.
KW Cell wall; Structural protein; Repeat; Signal.
FT SIGNAL 1 214 POTENTIAL.
FT CHAIN ? 214 GLYCINE-RICH CELL WALL STRUCTURAL
FT DOMAIN 82 158 PROTEIN 2.
FT DOMAIN 176 195 GLY-RICH.
FT


```

DB      161 GGGGGGGGGGGG 174

RESULT 9
HXB3_MOUSE
ID      HXB3_MOUSE      STANDARD:      PRT:      433 AA.
AC      P09026; P10285; 061680;
DT      01-NOV-1988 (Rel. 09, Created)
DT      01-OCT-1996 (Rel. 34, Last sequence update)
DT      15-JUL-1999 (Rel. 38, Last annotation update)
DE      HOMEBOX PROTEIN HOX-B3 (HOX-2.7) (MH-23).
GN      HOXB3 OR HOXB-3 OR HOX-2.7.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=92258392; PubMed=1582411;
RA      Sham M.H., Hunt P., Nonchev S., Papalopulu N., Graham A.,
RA      Boncinelli E., Krumlauf R.;
RT      "Analysis of the murine Hox-2.7 gene: conserved alternative
RT      transcripts with differential distributions in the nervous system and
RT      the potential for shared regulatory regions.";
RL      EMBO J. 11:1825-1836(1992).
RN      [2]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=95196953; PubMed=7890121;
RA      Brown W.M., Taylor G.R.;
RT      "The 5'-sequence of the murine Hox-b3 (Hox-2.7) gene and its intron
RT      contain multiple transcription-regulatory elements.";
RL      Int. J. Biochem. 26:1403-1409(1994).
RN      [3]
RP      SEQUENCE OF 152-361 FROM N.A.
RX      MEDLINE=88054465; PubMed=2890503;
RA      Lonai P., Arman E., Czosnek H., Ruddle F.H., Blatt C.;
RT      "New murine homeoboxes: structure, chromosomal assignment, and
RT      differential expression in adult erythropoiesis.";
RL      DNA 6:409-418(1987).
RN      [4]
RP      SEQUENCE OF 181-265 FROM N.A.
RX      MEDLINE=89091992; PubMed=2463210;
RA      Graham A., Papalopulu N., Lortimer J., McVey J.H., Tuddenham E.G.D.,
RA      Krumlauf R.;
RT      "Characterization of a murine homeo box gene, Hox-2.6, related to the
RT      Drosophila Deformed gene.";
RL      Genes Dev. 2:1424-1438(1988).
RN      [5]
RP      FUNCTION: SEQUENCE-SPECIFIC TRANSCRIPTION FACTOR WHICH IS PART OF
RP      A DEVELOPMENTAL REGULATORY SYSTEM THAT PROVIDES CELLS WITH
RP      SPECIFIC POSITIONAL IDENTITIES ON THE ANTERIOR-POSTERIOR AXIS.
CC      -1- SIMILARITY: BELONGS TO THE ANTP FAMILY OF HOMEBOX PROTEINS.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X66177; CAA46951.1; -
DR      EMBL; 002278; AAB60496.1; -
DR      EMBL; M18168; AAA37840.1; -
DR      PIR; S20963; S20963.
DR      PIR; C29585; C29585.
DR      HSSP; P02833; ISAN.
DR      TRANSFAC; T01724; -
DR      MGD; MGI:96184; HOXB3.
DR      INTERPRO; IPR001356; -
DR      INTERPRO; IPR001827; -
DR      Pfam; PF00046; homeobox.1.
DR      PRINTS; PR00024; HOMEBOX.

```

```

DR      PRINTS; PR00025; ANTENNAPEDIA.
DR      PROSITE; PS00027; HOMEBOX.1; 1.
DR      PROSITE; PS00032; ANTENNAPEDIA; 1.
DR      PROSITE; PS00071; HOMEBOX.2; 1.
KW      Homeobox; DNA-binding; Developmental protein; Nuclear protein;
KW      Transcription regulation.
FT      DOMAIN 129 134
FT      DOMAIN 154 181
FT      DNA_BIND 191 250
FT      CONFLICT 113 113
FT      CONFLICT 119 119
FT      CONFLICT 152 169
FT      CONFLICT 182 182
FT      CONFLICT 216 217
FT      CONFLICT 330 330
FT      CONFLICT 342 361
FT      SEQUENCE 433 AA; 44353 MW; 9ADB3C922663612A6 CRC64;

Query Match      77.8%; Score 63; DB 1; Length 433;
Best Local Similarity 78.6%; Pred No. 2.5;
Matches 11; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy      1 GGGGGGGGRRASGGGG 14
Db      162 GGGGGGGGGSGGGG 175

RESULT 10
ID      KDFE_DROME      STANDARD:      PRT:      1454 AA.
AC      009103;
DT      01-FEB-1995 (Rel. 31, Created)
DT      01-FEB-1995 (Rel. 31, Last sequence update)
DT      15-JUL-1999 (Rel. 38, Last annotation update)
DE      EYE-SPECIFIC DIACYLGLYCEROL KINASE (EC 2.7.1.107) (RETINAL
DE      DEGENERATION A PROTEIN) (DIGLYCERIDE KINASE) (DGK) (DAG KINASE)..
GN      RDGA OR DGR2.
OS      Drosophila melanogaster (Fruit fly).
OC      Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC      Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC      Ephydroidea; Drosophilidae; Drosophila.
RN      [1]
RP      SEQUENCE FROM N.A.
RX      STRAIN-CANTON-S; TISSUE-HEAD;
RX      MEDLINE=94068563; PubMed=8248222;
RA      Masai I., Okazaki A., Hosoya T., Hotta Y.;
RT      "Drosophila retinal degeneration A gene encodes an eye-specific
RT      diacylglycerol kinase with cysteine-rich zinc-finger motifs and
RT      ankyrin repeats.";
RL      Proc. Natl. Acad. Sci. U.S.A. 90:11157-11161(1993).
RN      [2]
RP      FUNCTION: REQUIRED FOR THE MAINTENANCE OF THE PHOTORECEPTOR. ITS
RP      ABSENCE LEADS TO RHABDOMERE DEGENERATION DUE TO DEFECTIVE
RP      PHOSPHOLIPID TURNOVER.
CC      -1- CATALYTIC ACTIVITY: ATP + 1,2-DIACYLGLYCEROL = ADP +
CC      1,2-DIACYLGLYCEROL 3-PHOSPHATE.
CC      -1- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED.
CC      -1- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN ADULT EYE.
CC      -1- DISEASE: MUTANTS OF THIS GENE ARE CHARACTERIZED BY HAVING
CC      PHOTORECEPTOR CELLS THAT DEVELOP NORMALLY BUT DEGENERATE RAPIDLY
CC      AFTER ECDIOSION.
CC      -1- SIMILARITY: BELONGS TO THE EUKARYOTIC DIACYLGLYCEROL KINASE
CC      FAMILY.
CC      -1- SIMILARITY: CONTAINS 2 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
CC      BINDING DOMAINS.
CC      -1- SIMILARITY: CONTAINS 4 ANK REPEATS.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way

```

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: D17315; BAA04135.1; -
 DR FLYBASE; FBgn0003217; rdga.
 DR INTERPRO; IPR000756; -
 DR INTERPRO; IPR001206; -
 DR INTERPRO; IPR002110; -
 DR INTERPRO; IPR002219; -
 DR PFAM; PF00609; DAGKc; 1.
 DR PFAM; PF00781; DAGKc; 1.
 DR PFAM; PF00130; DAG-PE-bind; 1.
 DR PFAM; PF00023; ank; 3.
 DR PROSITE; PS50088; ANK_REPEAT; 2.
 DR PROSITE; PS50297; ANK_REPEAT; 1.
 DR PROSITE; PS50479; DAG-PE_BIND_DOM_1; FALSE_NEG.
 DR PROSITE; PS50081; DAG-PE_BIND_DOM_2; FALSE_NEG.
 DR TRANSFERASE; Kinase; ANK repeat; Repeat; Vision;
 KW Phorbol-ester binding.
 FT DOMAIN 2 6
 FT DOMAIN 24 39
 FT DOMAIN 110 115
 FT DOMAIN 227 231
 FT DOMAIN 758 775
 FT DOMAIN 592 642
 FT DOMAIN 662 719
 FT DOMAIN 807 935
 FT DOMAIN 961 1115
 FT DOMAIN 1317 1451
 FT DOMAIN 1317 1349
 FT REPEAT 1350 1382
 FT REPEAT 1386 1418
 FT REPEAT 1419 1451
 FT MUTAGEN 869 869 G->D: IN MUTANT RDC42.
 FT SEQUENCE 1454 AA; 159675 MW; CE4C81099FEA16AA CRC64;

Query Match 77.8%; Score 63; DB 1; Length 1454;
 Best Local Similarity 78.6%; Pred. No. 6.9;
 Matches 11; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGGGSGGRASGGG 14
 DB 760 GCGGAGGAGGAGGGG 773

RESULT 11
 GRP_MAIZE STANDARD; PRT; 157 AA.
 AC P10979;
 DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DT 01-FEB-1995 (Rel. 31, Last annotation update)
 DE GLYCINE-RICH RNA-BINDING, ABSICISIC ACID-INDUCIBLE PROTEIN.
 GN RAB15.
 OS Zea mays (Maize).
 RT Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.
 RN 11
 RP SEQUENCE FROM N.A.
 RC TISSUE-ENDOSPERM;
 RX MEDLINE-88288401; PubMed-2969461;
 RA Gomez J., Sanchez-Martinez D., Stiefel V., Rigau J., Puigdomenech P.,
 RA Pages M.;
 RA "A gene induced by the plant hormone abscisic acid in response to
 RT water stress encodes a glycine-rich protein.";
 RL Nature 334:262-264 (1988).
 RN 12
 RP SIMILARITY TO RNA-BINDING PROTEINS.
 RX MEDLINE-89097302; PubMed-2521378;

RA Mortenson E., Dreyfuss G.;
 RA "RNP in maize protein.";
 RL Nature 337:312-312(1989).
 CC -1- FUNCTION: POSSIBLY HAS A ROLE IN RNA TRANSCRIPTION OR
 CC PROCESSING DURING STRESS.
 CC -1- INDUCTION: BY THE PLANT HORMONE ABSICISIC ACID IN RESPONSE TO
 CC WATER STRESS.
 CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RNM).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X12564; CAA31077.1; -
 DR PIR: S04536; S04536.
 DR HSSP; P09651; 1UP1.
 DR MAIZEDB; 69261; -
 DR INTERPRO; IPR000504; -
 DR PFAM; PF00076; rrm; 1.
 DR PROSITE; PS50102; RRM; 1.
 DR PROSITE; PS00030; RRM_RNP_1; 1.
 KW RNA-binding.
 FT DOMAIN 10 15
 FT DOMAIN 49 56
 FT DOMAIN 88 154
 FT SEQUENCE 157 AA; 15438 MW; 8A10592248B60D16 CRC64;

Query Match 76.5%; Score 62; DB 1; Length 157;
 Best Local Similarity 78.6%; Pred. No. 1.3;
 Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGGGSGGRASGGG 14
 DB 116 GCGGCGGRREGGG 129

RESULT 12
 GRP_DAUCA STANDARD; PRT; 157 AA.
 AC Q03878;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE GLYCINE-RICH RNA-BINDING PROTEIN.
 OS Daucus carota (Carrot).
 OS Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Asteridae;
 OC easterids II; Apiales; Apiaceae; Daucus.
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. QUEEN ANNE'S LACE;
 RA Sturm A.;
 RT "A wound-inducible glycine-rich protein from Daucus carota with
 RT homology to single-stranded nucleic acid binding proteins.";
 RL Plant Physiol. 99:1689-1692(1992).
 CC -1- FUNCTION: MAY PLAY A ROLE IN THE BIOSYNTHESIS AND PROCESSING OF
 CC HETEROGENEOUS NUCLEAR RNA AND IN THE MATURATION OF SPECIFIC MNAS
 CC IN RESPONSE TO WOUNDING.
 CC -1- INDUCTION: IN RESPONSE TO STRESS BY WOUNDING.
 CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RNM).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

```

CC -----
DR EMBL: X58146; CAA41152.1; -
DR PIR: S14857; S14857.
DR HSSP: P09651; 10P1.
DR INTERPRO: IPR000504; -.
DR PFAM: PF00076; rrm; 1.
DR PROSITE: PS0102; RRM; 1.
DR PROSITE: PS00030; RRM_RNP_1; 1.
DR RNA-binding.
KW DOMAIN 8 13 RNA-BINDING (RNP2) (BY SIMILARITY).
FT DOMAIN 47 54 RNA-BINDING (RNP1) (BY SIMILARITY).
FT DOMAIN 86 154 GLY-RICH.
SO SEQUENCE 157 AA; 15718 MW; 73FBD644F51CB633 CRC64;

Query Match
Best Local Similarity 76.5%; Score 62; DB 1; Length 157;
Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGGGSGGRASGGG 14
    ||||| ||| ||||
Db 103 GGGGSGGRGREGGG 116

RESULT 13
GRP1_ORYSA STANDARD; PRT; 165 AA.
ID P25074;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE GLYCINE-RICH CELL WALL STRUCTURAL PROTEIN 1 PRECURSOR.
GN GRP-1.
OS Oryza sativa (Rice).
OC Eurytyota; Viridiplantae; Embryophyta; Tracheophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. INDICA-IR36;
RX MEDLINE-91370862; PubMed-1716496;
RA "Lei M., Wu R.;
RL "A novel glycine-rich cell wall protein gene in rice."
RL Plant Mol. Biol. 16:187-198(1991).
CC -1- FUNCTION: RESPONSIBLE FOR PLASTICITY OF THE CELL WALL (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: CELL WALL (POTENTIAL).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X53596; CAA37665.1; -
DR PIR: S13385; KNRZG1.
DR HSSP: P30129; 4DPV.
KW Cell wall; Structural protein; Repeat; Signal.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 165 GLYCINE-RICH CELL WALL STRUCTURAL
FT PROTEIN 1.
FT DOMAIN 31 159 GLY-RICH.
FT REPEAT 56 62 R2 (TYR-RICH).
FT REPEAT 93 99 R2 (TYR-RICH).
FT REPEAT 132 138 R2 (TYR-RICH).
SO SEQUENCE 165 AA; 13536 MW; E36CE31C3650AC9A CRC64;

Query Match
Best Local Similarity 76.5%; Score 62; DB 1; Length 165;
Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GGGGSGGRASGGG 15

```

```

Db 111 GGGGSGGGGGGGG 125
    ||||| ||| |||||

RESULT 14
GRP2_SORBI STANDARD; PRT; 168 AA.
ID GRP2_SORBI 099070;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE GLYCINE-RICH RNA-BINDING PROTEIN 2.
GN GRP2.
OS Sorghum bicolor (Sorghum) (Sorghum vulgare).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Sorghum.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. TAMARAN FNK 140; TISSUE-LEAF;
RX MEDLINE-91346715; PubMed-1715211;
RA Crelin C., Puigdomenech P.;
RT "Glycine-rich RNA-binding proteins from Sorghum vulgare."
RL Plant Mol. Biol. 15:783-785(1990).
CC -1- FUNCTION: POSSIBLY HAS A ROLE IN RNA TRANSCRIPTION OR
CC PROCESSING DURING STRESS.
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X57662; CAA40862.1; -
DR PIR: S12312; S12312.
DR HSSP: P09651; 10P1.
DR INTERPRO: IPR000504; -.
DR PFAM: PF00076; rrm; 1.
DR PROSITE: PS0102; RRM; 1.
DR PROSITE: PS00030; RRM_RNP_1; 1.
KW RNA-binding.
FT DOMAIN 10 15 RNA-BINDING (RNP2) (BY SIMILARITY).
FT DOMAIN 49 56 RNA-BINDING (RNP1) (BY SIMILARITY).
FT DOMAIN 88 157 GLY-RICH.
SO SEQUENCE 168 AA; 16360 MW; 92F4FD07FE4145D8 CRC64;

Query Match
Best Local Similarity 76.5%; Score 62; DB 1; Length 168;
Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGGGSGGRASGGG 14
    ||||| ||| ||||
Db 114 GGGGSGGRGREGGG 127

RESULT 15
CHIA_MAIZE STANDARD; PRT; 280 AA.
ID CHIA_MAIZE P29022;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE ENDOCHITINASE A PRECURSOR (EC 3.2.1.14) (SEED CHITINASE A).
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta;
OC Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-SEED;
RX MEDLINE-92202208; PubMed-1551872;

```


